UNITED STATES DEPARTMENT OF COMMERCE
Bureau of the Census
Washington, DC 20233-0001

September 21, 2000

DSSD CENSUS 2000 PROCEDURES AND OPERATIONS MEMORANDUM SERIES B-8

MEMORANDUM FOR Howard Hogan
Chief, Decennial Statistical Studies Division

From: Donna Kostanich
Assistant Division Chief, Sampling and Estimation
Decennial Statistical Studies Division

Prepared by: Thomas Mule
Sample Design Team
Decennial Statistical Studies Division

Subject: Accuracy and Coverage Evaluation Survey: Decomposition of Dual System Estimate Components (Prototype)

The attached document is a prototype of the report that we will prepare, per your request, following completion of applicable Accuracy and Coverage Evaluation Survey (A.C.E.) operations. The completed report is intended to aid the Executive Steering Committee on A.C.E. Policy (ESCAP) in its recommendation regarding the release of the statistically corrected data or the data without statistical correction as the P.L. 94-171 data. This report, together with other reports, will assess the operations and results of both the initial Census and the A.C.E. Both sets of assessments will be available to the ESCAP to aid the Committee in reaching its recommendation regarding the use of the statistically corrected data.

This report focuses on decomposition of Dual System Estimate components. The analysis is limited to the outcome of the estimation steps on the components of the dual system estimates. Attached to this prototype document is an example table.

It is important to note that the conduct of the operations may lead us to modify the attached format by including additional information. It is also likely that descriptions and definitions will be enhanced or the data items could undergo revision. Conversely, we may conclude, for a variety of reasons, that some of the information set forth in the attached prototype may not be available. The attached document sets forth our conclusions prior to completion of the A.C.E. about what information would properly inform the ESCAP on this subject, but is subject to modification.

Accuracy and Coverage Evaluation 2000: Decomposition of Dual System Estimate Components

prepared by Thomas Mule

Introduction

Before dual system estimates, a number of estimation steps occur. The purpose of this document is to show the outcome of the estimation steps on the dual system estimates (DSEs) by showing how these steps contribute to the estimated components of dual system estimation.

This document examines the outcome of the following estimation steps:

- Noninterview adjustments for whole-household nonresponse in the P sample,
- Characteristic imputation of missing post-stratification variables for the P and E samples,
- Imputation Cell Estimation (ICE) which assigned probabilities to people with unresolved match and residence status for the P sample and unresolved enumeration status for the E sample,
- Targeted Extended Search (TES) which expanded the search area for selected cases during person matching to reduce the variance of the DSEs.

This report documents the decomposition of the DSE components for the 2000 A.C.E. The results, provided in the attached tables, show the outcome of the estimation steps on the DSE component estimates. The text of this report addresses the estimation steps, DSE components and a description of the results.

Overall Assessment

Prototype Note: An overall assessment of the information provided in the attached tables will be provided here.

Estimation Steps

This section provides a description of each estimation step in the P sample and E sample examined in this analysis. See Cantwell (2000) for more information on the results of the estimation steps implemented during the Missing Data process. The last part of this section describes TES which applies to both samples.

P Sample

Noninterview Adjustments.

Procedure C uses in-movers along with non-movers to estimate the number of P-Sample people in the post-stratum, while using out-movers to estimate the match rate of movers in the post-stratum. Because of the use of Procedure C estimation, there are two noninterview adjustments. The **Census Day noninterview adjustment** is based on the housing-unit status as of Census Day (i.e., the Census Day roster), and the **A.C.E. Day noninterview adjustment** is based on the housing-unit status as of the day of the A.C.E. interview (i.e., the Interview Day roster).

Interview status for each adjustment is determined by the following. The reference day for the Census Day noninterview adjustment is Census day while the reference day for the A.C.E. Day noninterview adjustment is the A.C.E. interview day.

> Interview: A unit is an interview (for the given reference date) if there is at least one person (with name and at least two demographic characteristics) who possibly or definitely was a resident of the housing unit on the given reference date.
>
> Noninterview: An occupied housing unit (as of the given reference date) that is not an interview is a noninterview.
>
> Not an occupied housing unit: A housing unit determined to be vacant or a nonexistent housing unit on the reference day.

The **Census Day noninterview adjustment** is used to adjust the person weights of **nonmovers and outmovers**. The **A.C.E. Day noninterview adjustment** is used to adjust the person weights of **inmovers**.

Characteristic Imputation for Missing Post-stratification Variables

P-Sample characteristic imputation uses either a hot deck procedure or draws from available demographic distributions. The following variables, needed to assign people to post-strata, are imputed if missing:

- Age
- Race
- Hispanic Origin
- Sex
- Tenure

All of the variables are person-level except for tenure which is housing unit level. Tenure is imputed from the previous household with a similar type of basic address with tenure recorded.

For the person-level variables, both these procedures attempt to use within household information whenever possible. Attachment A provides more information on the characteristic imputation.

Residence Status ICE

The A.C.E. needs to determine if nonmovers and outmovers were residents on Census Day. People with unresolved residence status were collected in the person interview as nonmovers or outmovers, but not enough information was collected to resolve their residence status The residence status ICE assigns probabilities for P-Sample people with unresolved residence status. P-sample persons are separated into groups called imputation cells based on operational, demographic, or geographic characteristics. The weighted average of 1's and 0's (representing, e.g., residence and non-residence, respectively) is calculated for each cell, and that value is imputed for all persons in the cell with the unresolved status. Residence status is determined by the Final Match Code. Attachment B provides more information on the Final Match Codes for the P sample.

Match Status ICE

The match status ICE assigns probabilities for P-Sample people with unresolved match status. P-sample persons are separated into groups called imputation cells based on operational, demographic, or geographic characteristics. The weighted average of 1's and 0's (representing, e.g., match and non-match, respectively) is calculated for each cell, and that value is imputed for all persons in the cell with the unresolved status. Match status is determined by the Final Match Code. Attachment B provides more information on the Final Match Codes for the P sample.

E Sample

Characteristic Imputation for Missing Post-stratification Variables

Characteristic imputation for the E sample can be done in either of two processes. Below is a description of each process. Attachment A provides more information on characteristic imputation.

The following characteristic variables, needed to assign people to post-strata, are imputed if missing:

- Age
- Race
- Hispanic Origin
- Sex
- Tenure

Census Editing and Imputation

Since the E sample is a sample of enumerations, the first attempt is to use the results from Hundred Percent Census Edited File. The 100% Edits and Allocations is the process of editing inconsistent 100% data items and imputing for missing 100% data items collected during the census enumeration. The purpose of the allocations is to ensure that each person and household on the final hundred percent data file has a value for each of the 100% data items. In addition, the edits serve to ensure certain consistencies between characteristics. The population items that go through the processing include relationship, sex, race, origin, age and date of birth. Household items include householder and spouse determination and tenure.

Missing Data E-sample Characteristic Imputation

The Missing Data process has a backup imputation plan in case E-sample cases still having missing values after using the Hundred Percent Census Edited File. The backup plan is similar to the characteristic imputation for the P sample.

Correct Enumeration ICE

The Correct Enumeration ICE assigns probabilities for E-Sample people with unresolved enumeration status. E-sample persons are separated into groups called imputation cells based on operational, demographic, or geographic characteristics. The weighted average of 1's and 0's (representing, e.g., correct enumeration and erroneous enumeration, respectively) is calculated for each cell, and that value is imputed for all persons in the cell with the missing status. Enumeration status is determined by the Final Match Code. Attachment C provides more information on the Final Match Codes for the E sample.

TES

Both the E sample and the P sample measure enumeration errors in the census. The E sample measures gross overcount in the form of erroneous enumerations. The P sample measures gross undercount in the form of nonmatches. Ideally, the entire census could be searched before a P-sample person was declared to be a nonmatch. Likewise, the entire country could ideally be searched to determine if an E-sample enumeration is a duplicate or is fictitious. Of course, such extensive searches are not feasible. For this reason, determining the search area is key for a coverage evaluation survey.

For the 1990 Post-Enumeration Survey (PES), matching was first done within the block. Depending on the type of enumeration area, a surrounding block search was performed in each cluster by expanding the search area for both the P sample and E sample. The motivation for extending the search area is variance reduction. Allowing more cases to be matched and more census enumerations found in surrounding blocks result in a higher match and correct

enumeration rates and a DSE with more precision or less variance. Expanding the search area in every cluster was not efficient. In 1990, finding matches or duplicates in the surrounding block search was not a common event. There were many housing units in the surrounding blocks to search in order to match people and code census duplicates. For 2000, A.C.E. is targeting the expanded search to blocks containing geocoding errors in order to produce better quality data.

For 2000, the Bureau needed to develop an efficient operation of extending the search area that would produce variance reduction benefits. The default search area is the sample cluster which comprises one or more contiguous census collection blocks. In order to make the operation more efficient than 1990, TES focuses on people with certain errors. These people are called TES persons. The TES "targets" whole household geocoding errors in the E sample and whole household nonmatches in the P sample.

For TES, a P-sample TES person is one of the following:

- whole household person nonmatches in housing units that were nonmatches in the initial housing unit matching OR
- P-sample whole household non-matches whose housing unit was matched in the initial housing unit phase to a census unit that was removed from the sample cluster.

For TES, an E-sample TES person is one of the following:

- persons in housing units identified in the housing unit match to be geocoding error OR
- census whole household non-matches in housing units added to the census files since initial housing unit matching and that were determined as a result of person followup to be geocoding error.

Since expanding the search area for each cluster was not efficient, this led to Sample Based Targeted Extended Search. We determined for efficiency reasons to extend the search area for 20 percent of the clusters. Olson (1999) researched various plans for selecting the clusters for TES. His research showed that geocoding error was highly clustered in the 1990 PES. About 72 percent of the census geocoding errors were found in less than 3 percent of the sample block clusters. The A.C.E. TES sampling selection method was a combination of certainty and probability sampling.

Certainty Sample

- Five percent of clusters with the most census geocoding errors and independent listing address nonmatches in the Initial Housing Unit phase
- Five percent of clusters with the most weighted census geocoding errors and A.C.E. address nonmatches in the Initial Housing Unit phase
- All relisted clusters in the P sample.

Probability sample

- 10 percent of the clusters selected by a systematic sample from the remaining clusters with at least one census geocoding error or an A.C.E. unmatched address based on the Initial Housing Unit phase. A.C.E sample clusters in List/Enumerate areas are out of scope for TES sample selection. Implementing TES in List/Enumerate clusters will be determined during Person Matching.

In the sample clusters, the search area for TES persons in the P and E samples is extended by one ring of census blocks. Wolfgang (1999) recommended extending the search area by only one ring. His research using Census 2000 Dress Rehearsal data showed that the additional benefits of using two rings of surrounding blocks were almost negligible.

The TES persons in the sample clusters have their final person weights adjusted to account for the TES sampling. The TES persons in non-selected clusters have their final person weights set equal to 0. These non-selected cases are represented by the extended search of the TES people in the sample clusters. See Childers (2000) for more information on TES.

The TES can introduce a bias called balancing error if the search area for the P-sample nonmatches is not the same as the search area for the E-sample correct enumerations.

Components of the DSE

Table 1 lists the components of the DSE examined in this report. This section describes each component. Attachment D provides more information and formulas for calculating the components in this analysis.

Table1 : DSE components

Sample	Component
E sample	Correct Enumerations (CE)
	Enumerations (N_e)
P sample	Non-movers (N_n)
	In-movers (N_i)
	Out-movers (N_o)
	Nonmover matches (M_n)
	Outmover matches (M_o)
	Total Matches (M)
	Total People (N_p)
	Out-of-scope

Correct Enumerations (CE)

A correct enumeration is a person who is enumerated in a block cluster during the census who is also determined by A.C.E. operations to have lived in the search area on Census Day. These cases are people with a correct enumeration probability greater than zero; that is, correct enumerations consist of confirmed residents and unresolved residents.

Enumerations (N_e)

Persons enumerated in E-sample housing units during the Census. Please see Childers (2000) for more information.

Nonmovers (N_n)

P-sample persons who lived in the A.C.E. housing unit on Census Day and on the day of the A.C.E. interview.

Inmovers (N_i)

P-sample persons who lived in the A.C.E. housing unit on the day of the A.C.E. interview, but lived elsewhere on Census Day.

Outmovers (N_o)

P-sample persons who lived in the A.C.E. housing unit on Census Day, but moved elsewhere before the A.C.E. interview.

Nonmover Matches (M_n)

P-sample nonmovers who matched to a census enumeration in the search area during Person Matching.

Outmover Matches (M_o)

P-sample outmovers who matched to a census enumeration in the search area during Person Matching.

Total Matches (M)

The estimated number of P-sample matches. For Procedure C, the estimated number of matches is the sum of the nonmover matches and the estimated number of mover matches based on the outmover match rate times the number of inmovers:

$$M = M_n + \frac{M_o}{N_o} \times N_i$$

For Procedure A, used when fewer than 10 outmovers in a post-stratum, the estimated number of matches is the sum of the estimated nonmover and outmover matches:

$$M = M_n + M_o$$

Procedure A or Procedure C is applied to individual post-strata. Therefore, for aggregates of post-strata, the weighted P-Sample match estimate is simply the sum of the post-stratum P-Sample match estimates.

Total P-Sample People (N_p)

The estimated population size from the P sample. For Procedure C, the population estimate is the sum of the estimated nonmovers and inmovers:

$$N_p = N_n + N_i$$

For Procedure A, used when fewer than 10 outmovers in a post-stratum, the population estimate is the sum of the estimated nonmovers and outmovers:

$$N_p = N_n + N_o$$

Procedure A or Procedure C is applied to individual post-strata. Therefore, for aggregates of post-strata, the weighted P-Sample person estimate is simply the sum of the post-stratum P-Sample person estimates.

Out-of-Scope

Out-of-scope cases are A.C.E. interviewed cases determined to be one of the following:

- person was fictitious,
- person did not live in the block cluster on census day,
- person lived in group quarters or had another residence where the person should have been counted on census day,
- person was interviewed in a housing unit that exists outside the sample block cluster,
- person was a duplicate of another P-sample person

These cases are out-of-scope and do not contribute to the components of the dual system estimates.

Results

The results of the analysis leading up to the dual system estimates are provided in the attached tables. Each table analyzes the outcome of the estimation steps on the components of the dual system estimates. These results focus on the A.C.E. estimation for the 50 states and the District of Columbia.

A standard table shell format has been created to show the results of this analysis. **An example table has been attached to help explain the analysis of these tables.**

Levels of Analysis

We will aggregate the data to various levels and place those results in tables. The tables are organized as follows:

Attachment	Tables	Description
E	E1	2000 A.C.E. National results
F	F1 - F7	2000 A.C.E. results for the Race/Hispanic Origin domains
G	G1-G7	2000 A.C.E. results for the Age/Sex categories
H	H1-H2	2000 A.C.E. results for Owners and Non-Owners
I	I1-I14	2000 A.C.E. results for Race/Hispanic Origin domains by Owners and Non-Owners
J	J1-J448	2000 A.C.E. results for the 448 Post-strata

The results for the individual post-strata, Attachment J, will not be attached to this document due to its size but will be available electronically and provided upon request.

The race/Hispanic origin domains are:

- Domain 1 (American Indian or Alaska Native on reservations)
- Domain 2 (American Indian or Alaska Native off reservations)
- Domain 3 (Hispanic)
- Domain 4 (Non-Hispanic Black)
- Domain 5 (Native Hawaiian or Pacific Islander)
- Domain 6 (Non-Hispanic Asian)
- Domain 7 (Non-Hispanic White or "Some Other Race")

Describing the Tables

In order to show the outcome of the estimation steps on the components of the DSEs, we will describe the attached example table. This explanation will step through the rows in the table. For each row, we will identify the outcome being examined and how the outcome is being quantified.

Columns

The table columns are broken into 3 sections: E sample, P sample and rates.

In the E-sample section, the number of cases is given. There are columns for correct enumerations (CE), enumerations (N_e) and Remaining cases (Remain). In the P-sample section, the number of cases is given. There are columns for nonmover matches (M_n), outmover matches (M_o), nonmovers (N_n), inmovers (N_i), outmovers (N_o), Remaining cases (Remain), out-of-scope cases, total matches (M) and Total P-sample people (N_p). Most of these columns were described in the Components of the DSE section. The remaining cases (Remain) column for each sample identifies the number of cases that still need to be assigned in subsequent rows.

Summing the appropriate values down the column will equal the total at the bottom of the table for all but three columns. Two of the columns, the E-sample remaining cases and the P-sample remaining cases, identify the number of cases that still need to be assigned in subsequent rows. The third is the total matches column for the P sample. See Attachment D for a detailed explanation.

For each of these columns, three numbers are reported. Attachment D describes how to calculate the following three totals for each component.

- Sample size: number of cases for the component
- Unweighted: An unweighted estimate of the component. While weights are not used, the residence, match and correct enumeration probabilities are used where appropriate.
- Weighted: A weighted estimate of the component.

In the rates section, the correct enumeration rate, match rate and correction ratio are calculated. The correct enumeration rate is the correct enumerations (CE) estimate divided by the enumerations (N_e) estimate. The match rate is the total match (M) estimate divided by total P-sample people (N_p) estimate. The correction ratio is the correct enumeration rate divided by the match rate. This is NOT equal to the Coverage Correction Factor. Both unweighted and weighted estimates are calculated for each.

Rows

This part describes each row in the table. Each row quantifies a certain type of outcome on the dual system estimates. In the following, the cases corresponding to each type of outcomes are described. Information is provided on the weights used to calculate the weighted estimates. The description of each row ends by highlighting points from the example table.

Row 1: Resolved Cases Prior to Missing Data Adjustments (Non-TES)

Row 1 is the baseline to compare the outcome of the estimation steps. This row identifies the number of cases that meet all of the following criteria:

- no missing post-stratification variables,
- have resolved residence and match status for P-sample cases or have resolved enumeration status for E-sample cases,
- non-TES persons.

For weighted estimates, this row removes the noninterview adjustment from the final person weight for the P sample. For the E sample, there is no non-interview adjustment so the final person weight is used. This row allows us to examine the components of the DSE if the non-interview adjustment had not been applied to these cases.

Looking at the example table of the 1560 E-sample cases, 1333 are non-TES resolved cases. Of the 1333, 1250 were determined to be correct enumerations. This leaves 227 cases to be assigned to the subsequent rows. There are 1,809 cases on the Person DSE P-sample Person file that correspond to this table. Of these cases, 1586 are non-TES resolved nonmovers, 48 are non-TES resolved inmovers and 14 are non-TES resolved outmovers. The number of out-of-scope cases was 60. There are 101 cases remaining who need to be assigned to subsequent rows. Of the 1566 nonmovers, 1321 cases matched. Of the 14 outmovers, 8 matched. Using the Procedure C formula, the total matches (M) and total people (N_p) were estimated. A correct enumeration rate, match rate and correction ratio were calculated using the data from this row.

Row 2: Resolved Cases
Added by A.C.E. Day Noninterview Adjustment (Non-TES)

This row shows the outcome of the A.C.E. Day noninterview adjustment on the same cases identified in row 1, so sample sizes and unweighted estimates are not included. This row quantifies the weighted number of cases added to each component from the A.C.E. Day noninterview adjustment. This adjustment applies only to inmover cases in the P sample. Non-movers and out-movers in the P sample are not affected by this adjustment nor are the E-sample components.

Looking at the example table, 13.57 weighted inmovers were added by the noninterview adjustment. Using the Procedure C formula, N_p is increased by 13.57. If Procedure A were being used for this example, then N_p would not be affected by this adjustment.

Row 3: Resolved Cases
Added by Census Day Noninterview Adjustment (Non-TES)

This row shows the outcome of the Census Day noninterview adjustment on the same cases identified in row 1 so sample sizes and unweighted estimates are not included. This row quantifies the weighted number of cases added to each component from the Census Day non-interview adjustment. This adjustment applies only to non-mover and out-mover cases in the P

sample. Inmovers in the P sample are not affected by this adjustment nor are the E-sample components.

Looking at the example table, 649.42 weighted nonmovers and 9.38 weighted outmovers were added by the noninterview adjustment. A total of 538.11 weighted nonmover matches and 6.55 weighted outmover matches were also added.

Row 4: Characteristic Imputation Only (Non-TES)

This row shows the outcome of cases with all of the following criteria:

- imputed at least one missing post-stratification variable,
- resolved residence status and match status during Person Matching if in P sample
  OR
  resolved enumeration status during Person Matching if in E sample,
- non-TES person.

For weighted estimates, this row uses the final person weight for both P-sample and E-sample cases. The final weight for the P-sample cases includes the appropriate noninterview adjustment. Both P-sample and E-sample components can be affected by this outcome.

Looking at the example table, row 1 shows 227 cases remaining for the E sample. A total of 194 of the 227 needed characteristic imputation only and were non-TES cases. Of the 194 cases, 167 were correct enumerations. There are 33 cases remaining to be assigned to subsequent rows. For the P sample, row 1 shows 101 cases remaining. There are 46 nonmovers, two inmovers and five outmovers who needed characteristic imputation only and were non-TES cases. Of the 46 nonmover cases, 35 were matches. Of the five outmover cases, three were matches. Using the Procedure C formula, total match (M) and total person (N_p) estimates were calculated using the data from this row.

Row 5: Characteristic Imputation and Residence ICE (Non-TES)

This row shows the outcome of cases with all of the following criteria:

- imputed at least one missing post-stratification variable,
- unresolved residence status,
- resolved match status,
- non-TES persons.

Only P-sample cases are in this row because residence and match status only apply to the P sample. Only nonmovers and outmovers can meet this row's criteria. Inmovers have a resolved residence status and are not matched during Person Matching. Thus, the P-sample components are affected by this row while the E-sample components are not.

For weighted estimates, this row uses the final person weight for the P-sample cases. The final weight for the P-sample cases includes the appropriate noninterview adjustment.

Looking at the example table, there was one nonmover case that met the criteria of this row. This case is shown in the nonmover (N_n) column. The unweighted total of 0.83 shows the probability assigned to this case by Residence ICE. The weighted total of 6.52 shows the product of the residence probability and the final person weight. This nonmover record did not match so the nonmover matches (M_n) column is zero.

Row 6: Residence ICE Only (Non-TES)

This row shows the outcome of cases with all of the following criteria:

- no missing post-stratification variables,
- unresolved residence status,
- resolved match status,
- non-TES persons.

Only P-sample cases are in this row because residence and match status only apply to the P sample. Only nonmovers and outmovers can meet this row's criteria. Inmovers have a resolved residence status and are not matched during Person Matching. Thus, the P-sample components are affected by this row while the E-sample components are not.

For weighted estimates, this row uses the final person weight for the P-sample cases. The final weight for the P-sample cases includes the appropriate noninterview adjustment.

Looking at the example table, there were three nonmovers and one outmover that met this row's criteria. The unweighted and weighted totals in the nonmover (N_n) and outmover (N_o) columns are calculated the same as row 5. None of these cases matched so the nonmover matches (M_n) and outmover matches (M_o) columns are zero.

Row 7: Characteristic Imputation, Residence ICE and Match ICE (Non-TES)

This row shows the outcome of cases with all of the following criteria:

- imputed at least one missing post-stratification variable,
- unresolved residence status,
- unresolved match status,
- non-TES persons

Only P-sample cases are in this row because residence and match status only apply to the P sample. Only nonmovers and outmovers can meet this row's criteria. Inmovers have a resolved

residence status and are not matched. Thus, the P-sample components are affected by this row while the E-sample components are not.

For weighted estimates, this row uses the final person weight for the P-sample cases. The final weight for the P-sample cases includes the appropriate noninterview adjustment.

Looking at the example table, a total of 20 of the 43 remaining P-sample cases met this row's criteria. There are 17 nonmovers (N_n) and three outmovers (N_o). The unweighted totals for N_n and N_o sum each record's residence probability. The weighted totals for N_n and N_o sum the product of each record's residence probability and final person weight. The table shows a sample size of 17 nonmover matches (M_n) and three outmover matches (M_o). These are the number of cases with a match probability greater than 0. The unweighted totals for the M_n and M_o columns sum the product of each record's residence probability and the match probability. In this row, there are 12.84 unweighted nonmover matches and 2.29 unweighted outmover matches. The weighted totals sum the product of the each record's residence probability, match probability and final person weight. There are 149.93 weighted nonmover matches and 34.65 weighted outmover matches in this row.

Row 8 : Residence ICE and Match ICE (Non-TES)

This row shows the outcome of cases with all of the following criteria:

- no missing post-stratification variables,
- unresolved residence status,
- unresolved match status,
- non-TES persons

Only P-sample cases are in this row because residence and match status only apply to the P sample. Only nonmovers and outmovers can meet this row's criteria. Inmovers have a resolved residence status and are not matched. Thus, the P-sample components are affected by this row while the E-sample components are not.

For weighted estimates, this row uses the final person weight for the P-sample cases. The final weight for the P-sample cases includes the appropriate noninterview adjustment.

Looking at the example table, a total of 13 of the 23 remaining P-sample cases met this row's criteria. There are 10 total nonmovers (N_n) and three total outmovers (N_o). The unweighted totals for N_n and N_o sum each record's residence probability. The weighted totals for N_n and N_o sum the product of each record's residence probability and the final person weight. The table shows a sample size of 10 nonmover matches (M_n) and three outmover matches (M_o). These are the number of cases with a match probability greater than 0. The unweighted totals for the M_n and M_o columns sums the product of the each record's residence probability and match probability. There are 7.62 unweighted nonmover matches and 2.29 unweighted outmover

matches in this row. The weighted totals sum the product of each record's residence probability, match probability and final person weight. There are 85.95 weighted nonmover matches and 19.93 weighted outmover matches in this row.

Row 9: Characteristic Imputation and Correct Enumeration ICE (Non-TES)

This row shows the outcome of cases with all of the following criteria:

- imputed at least one missing post-stratification variable,
- unresolved enumeration status,
- non-TES people

Only E-sample cases are in this row because enumeration status only applies to the E sample. Thus, the E-sample components are affected by this row while the P-sample components are not.

For weighted estimates, this row uses the final person weight for the E-sample cases. The E sample does not have a noninterview adjustment.

Looking at the example table, a total of eight of the 33 remaining E-sample cases meet this row's criteria. The tables shows a sample size of eight correct enumerations (CE). This is the number of cases with a correct enumeration probability greater than 0. The unweighted total of CEs is 6.6. The unweighted total sums each record's correct enumeration probability. The weighted total of CEs is 72.28. The weighted total sums the product of each record's correct enumeration probability and the final person weight. The table shows a sample size and unweighted total of eight enumerations (N_n). There are 87.81 weighted enumerations in this row. The weighted total sums just the final person weight.

Row 10: Correct Enumeration ICE Only (Non-TES)

This row shows the outcome of cases with the following criteria:

- no missing post-stratification variables,
- unresolved enumeration status,
- non-TES people

Only E-sample cases are in this row because enumeration status only applies to the E sample. Thus, the E-sample components are affected by this row while the P-sample components are not.

For weighted estimates, this row uses the final person weight for the E-sample cases. The E sample does not have a noninterview adjustment.

Looking at the example table, a total of 15 of the 25 remaining E-sample cases met this row's criteria. The tables shows a sample size of 15 correct enumerations (CE). This is the number of cases with a correct enumeration probability greater than 0. The unweighted total of CEs is

11.89. The unweighted total sums each record's correct enumeration probability. The weighted total of CEs is 123.58. The weighted total sums the product of each record's correct enumeration probability and the final person weight. The table shows a sample size and unweighted total of 15 enumerations (N_e). There are 156.73 weighted enumerations in this row. The weighted total sums just the final person weight.

Row 11: TES People

This row shows the outcome of the TES people on the estimates. TES people in the E and P samples were identified in a previous section. These cases may have needed characteristic imputation, residence ICE or match ICE. This row focuses on the overall outcome of TES on the components of the DSE. Both E-sample and P-sample components are affected by this row.

For weighted estimates, this row uses the final person weight for both the E-sample and P-sample cases. For the E sample, the final person weight accounts for the TES sampling. For the P sample, the final person weight accounts for the appropriate noninterview adjustment and TES sampling.

Table 2 shows the national correct enumeration rate and match rate by TES status. The rates are shown for non-TES cases, TES cases and overall. Weighted rates are presented in the table.

Table 2: National Rates by TES Status (Weighted)

	Correct Enumeration Rate	Match Rate
Non-TES cases		
TES Cases		
Overall		

Looking at the example table, the remaining 10 E-sample and 10 P-sample cases from row 10 are TES people. In the E sample, the correct enumeration (CE) sample size and unweighted total is 6. The unweighted total of correct enumerations sums each record's correct enumeration probability. The weighted total of correct enumerations is 73.20. The weighted total sums the product of the each record's correct enumeration probability and the final person weight. The enumeration (N_e) sample size and unweighted total of enumerations are 10. The weighted total of enumerations is 104.62.

In the P sample, all 10 are nonmovers. The sample size and unweighted total for nonmovers (N_n) is 10. The unweighted total sums the each record's residence probability. The weighted total of nonmovers is 89. The sample size and unweighted total of nonmover matches (M_n) is 6. Unweighted total of nonmover matches sums the product of each record's residence probability and match probability. The weighted total of nonmover matches is 62.3. The weighted total

sums the product of each record's residence probability, match probability and final person weight.

Row 12: Total

The last row presents the totals for the table. The table shows the cumulative sample size, unweighted total and weighted total for the following columns:

- Correct Enumerations (CE)
- Enumerations (N_e)
- Nonmover matches (M_n)
- Outmover matches (M_o)
- Nonmovers (N_n)
- Inmovers (N_i)
- Outmovers (N_o)
- Total Matches (M)
- Total People (N_p)
- Out-of-scope

Using the totals for the components, the correct enumeration rate, match rate and correction ratio are calculated. For each, we calculate both unweighted and weighted estimates.

The remaining columns for the P-sample and the E-sample are 0. All cases have been assigned to one of the 11 rows.

References

Cantwell (2000) DSSD Census 2000 Procedure and Operations Memorandum Series, Chapter B, "Accuracy and Coverage Evaluation: Missing Data Results (Prototype)" September 21, 2000.

Childers (2000) DSSD Census 2000 Procedures and Operations Memorandum Series, Chapter S-DT-01, "Accuracy and Coverage Evaluation: The Design Document," June 20, 2000.

Olson (1999) DSSD Census 2000 Procedures and Operations Memorandum Series, Chapter Q-13, "Accuracy and Coverage Evaluation Survey: Targeted Extended Search Empirical Results," October 1, 1999

Wolfgang (1999) DSSD Census 2000 Procedures and Operations Memorandum Series, Chapter S-RE-2,"Research on Surrounding Block Rings for the Census 2000 Accuracy and Coverage Evaluation Processing," February 4, 2000

Characteristic Imputation

P Sample

For dual system estimation, P-sample records are imputed during Missing Data processing because either the value is missing or an edit failure. Table A1 identifies what is considered an imputed value for the variables needed to assign P-sample records to post-strata. The values in the table are the imputation flags values for these variables. The flags can be found on the P-sample Person DSE Output File.

Table A1: Identifying Imputed Values for the P sample

P-sample Characteristic	Reported Values	Imputed Values
Age	1 = No imputation	2 = Imputation because of edit failure 3 = Imputation because of missing value
Race	1 = No imputation	2 = Imputation because of edit failure 3 = Imputation because of missing value
Hispanic Origin	1 = No imputation	2 = Imputation because of edit failure 3 = Imputation because of missing value
Sex	1 = No imputation	2 = Imputation because of edit failure 3 = Imputation because of missing value
Tenure	1 = No imputation	2 = Imputation because of edit failure 3 = Imputation because of missing value

E Sample

For dual system estimation, E-sample records are imputed by one of two steps. First, we try to obtain the demographic and tenure variables from the Hundred Percent Census Edited File (HCEF) for each record. We identify if the HCEF has done any editing or imputation for these records. If the record does not match to the HCEF then the Missing Data process has a backup imputation system to impute missing values. Table A2 identifies what is considered an imputed value for the variables needed to assign E-sample records to a post-stratum. The values in the table are the HCEF allocation flag values for these variables. The flags can be found on the E-sample Person DSE Output File.

Table A2: Identifying Imputed Values for the E sample

E-sample Characteristic	Reported Values	Imputed Values
Age	0 = Both Consistent 1 = Age Only 2 = Date of birth only	3 = Inconsistent age and date of birth 4 = Allocated from hot deck 9 = E-sample person did not match to the HCEF
Race	0 = As reported	3= Assigned from race response to Hispanic origin question 4 = Allocated from within household 5 = Allocated from hot deck 9 = E-sample person did not match to the HCEF
Hispanic Origin	0 = 1 reported origin 1 = 2 reported origin 2 = 3 reported origin	3 = Assigned Hispanic Origin from race code 4 = Allocated from within household 5 = Allocated from hot deck (surname used) 6 = Allocated from hot deck (surname not used) 9 = E-sample person did not match to the HCEF
Sex	0 = As reported	1 = From first name 4 = Allocated from hot deck 5 = Allocated from consistency check 9 = E-sample person did not match to the HCEF
Tenure	0 = As reported	1 = Assigned by consistency check 4 = Allocated from hot deck 9 = E-sample person did not match to the HCEF

A.C.E. P-sample Person Match Codes

Matched, Confirmed Resident

M	=	The P-sample and the census people were matched
MR	=	The P-sample follow-up interview determined that the matched person with unresolved residence status is a resident as of census day.

Matched, Unresolved Residence Status

MU	=	The A.C.E. person follow-up interview obtained no useful information to resolve the residence status for the matched person who had a residence status of unresolved before follow-up. The P-sample person's residence status is unresolved.

Not Matched, Confirmed Resident

NP	=	The P-sample person is not matched to an E-sample person. The P-sample person is considered to be a resident of census day.
NC	=	The P-sample nonmatch was found on the census roster. This person was not matched because only name was collected in the census for this person In a large household and the census person was not data defined. The P-sample person is considered to be a resident on census day.
NR	=	The P-sample person is identified as a resident in the block cluster on census day during the A.C.E. person follow-up interview.

Not Matched -- Unresolved Residence Status

NU	=	Not enough information is collected during the A.C.E. person follow-up interview to identify the P-sample person as a resident or nonresident in the block cluster. The match status for the P-sample person is nonmatch.

Unresolved Match and Residence Status

P	=	There is not enough information collected to determine if the possible match is a match or not. The match and residence status of the P-sample person are unresolved.
KI	=	Match not attempted for the P-sample person because the person has insufficient information for matching and follow-up.
KP	=	Match not attempted for the P-sample person.

Removed from the P-sample

FP	=	The P-sample person is fictitious in this block cluster
NL	=	The P-sample person did not live at the sample address or in the block cluster on census day and was listed as a nonmover or outmover in error.
NN	=	The P-sample person is identified as a nonresident in the block cluster on census day during the A.C.E. person follow-up interview, because the person lived in group quarters or had another residence where the person should have been counted on census day according to census residence rules.
GP	=	The P-sample person is removed because the person interview was conducted at a housing unit that exists outside the sample block cluster.
DP	=	The P-sample person is a duplicate of another P-sample person.
MN	=	The A.C.E. person follow-up interview determined that the matched person with unresolved residence status is not a resident in this housing unit or in this block cluster.

Attachment C

A.C.E. E-sample Person Enumeration Codes

Correctly Enumerated

M	=	The P-sample and E-sample people were matched.
CE	=	The E-sample nonmatch is identified as correctly enumerated during the A.C.E. person follow-up interview.
MR	=	The A.C.E. person follow-up interview determined that the matched person with unresolved residence status is a resident.

Erroneously Enumerated

GE	=	The E-sample person is erroneously enumerated in this block cluster, because the census housing unit is a geocoding error. The E-sample person should have been enumerated elsewhere in the census.
EE	=	The E-sample nonmatch is identified as erroneously enumerated from the A.C.E. person follow-up interview.
FE	=	The E-sample nonmatch is determined to be fictitious in this block cluster during the follow-up interview. The person may have existed, but should not have been enumerated in the census within this block cluster.
DE	=	The E-sample person is a duplicate of another E-sample person or a duplicate of a census person in a surrounding block.
MN	=	The A.C.E. person follow-up interview determined that the matched person with unresolved residence status is not a resident in this housing unit or in the block cluster.
KE	=	Match not attempted for the E-sample person. The name is blank or incomplete or the name is complete but there are one or no person characteristics.

Unresolved

UE	=	Not enough information is collected during the A.C.E. person follow-up interview to identify the E-sample person as correctly or erroneously enumerated in the E-sample.
MU	=	The A.C.E. person follow-up interview obtained no useful information to resolve the residence status for the match person with unresolved residence status.
P	=	There is not enough information collected to determine if the possible match is a match or not.
GU	=	The geographic work for the targeted extended search is unresolved.

Formula for Estimating E-sample Components

This section shows the formula for calculating unweighted and weight estimates of the E-sample components for each row in the analysis tables.

The following subscript notation is used:

j: Person record
k: Row in table

Correct Enumerations (CE)

Sample Size: The number of cases with correct enumeration probability greater than 0.

Unweighted:

$$CE_{unweighted} = \sum_{j \in row\ k} CEPROB_j$$

where $CEPROB_j$ is the correct enumeration probability for the *j*th record.

Weighted:

$$CE_{weighted} = \sum_{j \in row\ k} WEIGHT_j \times CEPROB_j$$

where $WEIGHT_j$ is the E-sample Final Weight reflecting A.C.E. sample design and TES for the *j*th record.

Enumerations (N_e)

Sample Size: The number of E-sample cases.

Unweighted:

$$N_{e\ unweighted} = \sum_{j \in row\ k} 1$$

Weighted:

$$N_{e\ weighted} = \sum_{j \in row\ k} WEIGHT_j$$

Formula for Estimating P-sample Components

This section shows the formula for calculating unweighted and weight estimates of the P-sample components for each row in the analysis tables. For weighted estimates, a general formula is given for each component. The last part of this section details which person weight to use for each row in the table. The same subscript notation as the E-sample is used.

Nonmovers (N_n)

Sample Size: The number of nonmovers with residence probability greater than 0.

Unweighted:

$$N_{n\ unweighted} = \sum_{j \in nonmover\ \&\ j \in row\ k} RPROB_j$$

where $RPROB_j$ is the residence probability

Weighted:

$$N_{n\ weighted} = \sum_{j \in nonmover\ \&\ j \in row\ k} WEIGHT_j \times RPROB_j$$

Inmovers (N_i)

Sample Size: The number of inmovers in the row. Inmovers are residents of the cluster on A.C.E. Day so residence probability is not needed.

Unweighted: The number of inmovers in the row. Inmovers are residents of the cluster on A.C.E. Day so residence probability is not needed.

Weighted:

$$N_{i\ weighted} = \sum_{j \in inmover\ \&\ j \in row\ k} WEIGHT_j$$

Outmovers (N_o)

Sample Size: The number of outmovers with residence probability greater than 0.

Unweighted:

$$N_{o\ unweighted} = \sum_{j \in outmover\ \&\ j \in row\ k} RPROB_j$$

Weighted:

$$N_{o\ weighted} = \sum_{j \in outmover\ \&\ j \in row\ k} WEIGHT_j \times RPROB_j$$

Nonmover Matches (M_n)

Sample Size: The number of nonmovers with match probability greater than 0.

Unweighted:

$$M_{n\ unweighted} = \sum_{j \in nonmover\ \&\ j \in row\ k} RPROB_j \times MPROB_j$$

where $MPROB_j$ is the match probability for the jth record.

Weighted:

$$M_{n\ weighted} = \sum_{j \in nonmover\ \&\ j \in row\ k} WEIGHT_j \times RPROB_j \times MPROB_j$$

Outmover Matches (M_o)

Sample Size: The number of outmovers with match probability greater than 0.

Unweighted:

$$M_{o\ unweighted} = \sum_{j \in outmover\ \&\ j \in row\ k} RPROB_j \times MPROB_j$$

Weighted:

$$M_{o\ weighted} = \sum_{j \in outmover\ \&\ j \in row\ k} WEIGHT_j \times RPROB_j \times MPROB_j$$

Total Matches (M)

The formula used to estimate the total matches depends on the number of outmovers for the post-stratum.

If the number of outmovers is 10 or greater then total matches is calculated using the Procedure C formula.

Sample Size:

$$M_{sample\ size} = M_{n\ sample\ size} + \left(\frac{M_{o\ sample\ size}}{N_{o\ sample\ size}} \times N_{i\ sample\ size} \right)$$

Unweighted:

$$M_{unweighted} = M_{n\ unweighted} + \left(\frac{M_{o\ unweighted}}{N_{o\ unweighted}} \times N_{i\ unweighted} \right)$$

Weighted:

$$M_{weighted} = M_{n\ weighted} + \left(\frac{M_{o\ weighted}}{N_{o\ weighted}} \times N_{i\ weighted} \right)$$

If the number of outmovers is less than 10 then total matches is calculated using the Procedure A formula.

Sample Size:

$$M_{\text{sample size}} = M_{\text{n sample size}} + M_{\text{o sample size}}$$

Unweighted:

$$M_{\text{unweighted}} = M_{\text{n unweighted}} + M_{\text{o unweighted}}$$

Weighted:

$$M_{\text{weighted}} = M_{\text{n weighted}} + M_{\text{o weighted}}$$

The formulas above for Procedure A or Procedure C are applied to individual post-strata. Therefore, for aggregates of post-strata, the weighted P-Sample match estimate is simply the sum of the post-stratum P-Sample match estimates.

Total People (N_p)

The formula used to estimate the total matches depends on the number of outmovers for the post-stratum.

If the number of out-movers is 10 or greater than total people is estimated using the Procedure C formula.

Sample Size:

$$N_{\text{p sample size}} = N_{\text{n sample size}} + N_{\text{i sample size}}$$

Unweighted:

$$N_{\text{p unweighted}} = N_{\text{n unweighted}} + N_{\text{i unweighted}}$$

Weighted:

$$N_{\text{p weighted}} = N_{\text{n weighted}} + N_{\text{i weighted}}$$

If the number of outmovers is less than 10 than total people is estimated using the Procedure A formula.

Sample Size:

$$N_{p\ \text{sample size}} = N_{n\ \text{sample size}} + N_{o\ \text{sample size}}$$

Unweighted:

$$N_{p\ \text{weighted}} = N_{n\ \text{weighted}} + N_{o\ \text{weighted}}$$

Weighted:

$$N_{p\ \text{weighted}} = N_{n\ \text{weighted}} + N_{o\ \text{weighted}}$$

The formulas above for Procedure A or Procedure C are applied to individual post-strata. Therefore, for aggregates of post-strata, the weighted P-Sample person estimate is simply the sum of the post-stratum P-Sample person estimates.

Out-of-Scope

Sample Size: The number of out-of-scope cases

Unweighted:

$$\text{Out-of-Scope}_{\text{unweighted}} = \sum_{j \in \text{out-of-scope}\ \&\ j \in \text{row k}} 1$$

Weighted:

$$\text{Out-of-Scope}_{\text{weighted}} = \sum_{j \in \text{out-of-scope}\ \&\ j \in \text{Analysis Category}} \text{WEIGHT}_j$$

Person Weight (Weight_j)

For all rows except for the Resolved Cases rows (rows 1-3), the P-sample Final TES-Adjusted Weight is $Weight_j$ in the formulas.

For row 1: Resolved Cases Prior to Missing Data Adjustment (Non-TES), the trimmed cluster weight is used.

For row 2: Resolved cases added by A.C.E. Day noninterview adjustment (Non-TES) and row 3: Resolved cases added by Census Day noninterview adjustment (Non-TES), the weight used is:

Final Weight - Trimmed Weight

By using this as the weight, these two rows are able to show the outcome of the respective non-interview adjustments.

Why summing total matches across rows does not equal the total matches in the total row if Procedure C is used?

This occurs because the total matches calculation involves a ratio. The ratio calculated in each row and summed across rows does not equal the ratio calculated using the component totals. This is shown in the following formula.

$$\sum_{rows}\left[M_{n,row} + \left(\frac{M_{o,row}}{N_{o,row}}\right) N_{i,row} \right] \neq M_{n,total} + \left(\frac{M_{o,total}}{N_{o,total}}\right) N_{i,total}$$

If Procedure A is used, then the total matches summed across rows does equal the number of total matches in the total row. Procedure A estimation does not require the mover ratio found in Procedure C estimation.

$$\sum_{rows}\left[M_{n,row} + M_{o,row} \right] = M_{n,total} + M_{o,total}$$

Procedure A or Procedure C is applied to individual post-strata. Therefore, for tables showing aggregates of post-strata, the weighted P-Sample match estimate is simply the sum of the post-stratum P-Sample match estimates.

		Prototype Example	09/21/2000	E sample (1560 Cases)			P sample (1809 Cases)									CE	Match	Correction
				CE	Ne	Remain	Mn	Mo	Nn	Nl	No	Remain	Out-of-Scope	M	Np	Rate	Rate	Ratio
Non-TES	1	Resolved Cases Prior to Missing Adj	Sample Size	1250	1333	227	1321	8	1586	46	14	101	60	1348.43	1634			
			Unweighted	1250	1333	227	1321	8	1586	46	14	99.21	60	1348 43	1634	0 9377	0 8252	1.1363
			Weighted	12757.37	13607.58	2351 01	13467.96	80 24	16151.39	475 732	138 99	1057.79	610 13	13742 60	16627.12	0 9375	0 8265	1.1343
	2	Resolved Cases Added A C.E. Day NI	Weighted				0	0	0	13 57	0	0	0	0	13.57			
	3	Resolved Cases Added Census Day NI	Weighted				538 11	6 55	649 42	0	9.38	0	0	538 11	649 42			
	4	Characteristic Imputation Only	Sample Size	167	194	33	35	3	46	2	5	48	0	36.20	48			
			Unweighted	167	194	33	35	3	46	2	5	46 21		36 20	48			
			Weighted	1713 79	2001.85	349 16	375 66	33 40	485 28	15.77	60.87	495 87		384.31	501 05			
	5	Characteristic Imputation and Residence ICE	Sample Size			33	0	0	1	0	0	47	0	0	1			
			Unweighted			33			0.83			45.38			0.83			
			Weighted			349.16			6 52			469 35			6.52			
	6	Residence ICE Only	Sample Size			33	0	0	3	0	1	43	0	0	3			
			Unweighted			33			2.48		0 83	42.07			2 48			
			Weighted			349 16			21.66		6.52	461.17			21.66			
	7	Characistic Imputation, Residence ICE and Match ICE	Sample Size			33	17	3	17	0	3	23	0	17	17			
			Unweighted			33	12 84	2 29	16.45		2.93	22.69		12.84	16 45			
			Weighted			349.16	149.93	34 65	192.11		44.4	224.66		149 93	192.11			
	8	Residence ICE and Match ICE	Sample Size			33	10	3	10	0	3	10	0	10	10			
			Unweighted			33	7.62	2.29	9.76		2.93	10		7.62	9.76			
			Weighted			349 16	85.95	19.83	110.12		25.54	89		85.95	110 12			
	9	Characteristic Imputation and Correct Enumeration ICE	Sample Size	8	8	25	0	0	0	0	0	10	0	0	0			
			Unweighted	6.6	8	25						10						
			Weighted	72 28	87 81	261.35						89						
	10	Correct Enumeration ICE only	Sample Size	15	15	10	0	0	0	0	0	10	0	0	0			
			Unweighted	11.89	15	10						10						
			Weighted	123 58	156.73	104 62						89						
TES	11	TES Cases	Sample Size	6	10	0	6	0	10	0	0	0	0	6	10			
			Unweighted	6	10		6		10					6	10			
			Weighted	73 20	104.62		62 30		89					62 30	89			
Total	12	Total	Sample Size	1446	1560	0	1389	17	1673	50	26	0	60	1406	1723			
			Unweighted	1441.49	1560		1382 46	15.58	1671 52	50	25.69		60	1398 04	1721.52	0 9240	0 8121	1.1378
			Weighted	14740 22	15958 59		14679 91	174.77	17705 5	505 072	265.7		610.13	14854 68	18210.57	0.9237	0.8157	1.1323